

# The Mouse Genome Database: Genotypes, Phenotypes, and Models of Human Disease

**Carol J. Bult\***, **Janan T. Eppig**, **Judith A. Blake**, **James A. Kadin**, **Joel E. Richardson**  
and the **Mouse Genome Database Group<sup>†</sup>**

The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609 USA

Received September 21, 2012; Accepted October 19, 2012

## ABSTRACT

**The laboratory mouse is the premier animal model for studying human biology because all life stages can be accessed experimentally, a completely sequenced reference genome is publicly available and there exists a myriad of genomic tools for comparative and experimental research. In the current era of genome scale, data-driven biomedical research, the integration of genetic, genomic and biological data are essential for realizing the full potential of the mouse as an experimental model. The Mouse Genome Database (MGD; <http://www.informatics.jax.org>), the community model organism database for the laboratory mouse, is designed to facilitate the use of the laboratory mouse as a model system for understanding human biology and disease. To achieve this goal, MGD integrates genetic and genomic data related to the functional and phenotypic characterization of mouse genes and alleles and serves as a comprehensive catalog for mouse models of human disease. Recent enhancements to MGD include the addition of human ortholog details to mouse Gene Detail pages, the inclusion of microRNA knockouts to MGD's catalog of alleles and phenotypes, the addition of video clips to phenotype images, providing access to genotype and phenotype data associated with quantitative trait loci (QTL) and improvements to the layout and display of Gene Ontology annotations.**

## INTRODUCTION

The laboratory mouse is widely recognized as the premier animal model for investigating genetic and cellular systems relevant to human biology and disease. A large arsenal of experimental genetic tools is available for mouse, including unique inbred strains, a complete reference genome (and deep-sequencing data for 17 additional

inbred lines), extensive genome variation maps (e.g. Single Nucleotide Polymorphisms) and technologies for directly and specifically manipulating the mouse genome. An international effort to knockout all mouse genes has produced an ES cell line resource covering over 18 000 genes (1) and the phenotyping phase has begun (2). New resources for complex trait mapping including the Collaborative Cross and Diversity Outbred mice are beginning to emerge (3,4). In the arena of human genetics and genomics, exome sequencing and the quest for lower and lower cost genome sequences will change again the way we approach computational and experimental methods for understanding the biology of the genome. The mouse is essential for the functional analysis and annotation of rapidly emerging human genomes through comparative genomics.

Realizing the full power of the mouse as a model of human biology depends, in part, on integrating the diverse genetic, genomic and phenotypic data for the mouse in ways that promote experimental and translational research. The central objective of the Mouse Genome Database (MGD) is to provide an integrative and comparative bioinformatics resource that supports the effective translation of information from experimental mouse models to uncover the genetic basis of human diseases. MGD is the highly curated, community model organism database for the laboratory mouse providing web and programmatic access to a complete catalog of mouse genes and genome features integrated with functional annotations, a comprehensive catalog of mutant alleles, phenotype annotations, human disease model annotations, variation data and sequence data. MGD went online via the World Wide Web in 1994, unifying and harmonizing several different databases of genetic map and allele information for the laboratory mouse. MGD has evolved rapidly, re-tooling and enhancing the database to adapt to the multitude of new data types, developing and upgrading data access tools for an increasingly diverse community of researchers, and adopting new database and software technologies as they have emerged and matured.

\*To whom correspondence should be addressed. Tel: +1 207 288 6324; Fax: +1 207 288 6132; Email: carol.bult@jax.org

<sup>†</sup>The members of the Mouse Genome Database Group are provided in the Acknowledgements.

MGD is the central component of a number of coordinated genome informatics projects that are part of the Mouse Genome Informatics (MGI) consortium (<http://www.informatics.jax.org>). Other database resources available through the MGI web portal include the Gene Expression Database (GXD) (5), the Mouse Tumor Biology Database (6), the Gene Ontology (GO) project (7) and the MouseCyc database of biochemical pathways (8). Taken together, these resources provide a combination of data breadth, depth, integration and quality that exists nowhere else for mouse.

## IMPROVEMENTS

The curation efforts within MGD focus on maintaining a catalog of genes and other genome features, functional annotation of mouse genes using Gene Ontology terms, annotation of phenotypes associated with genotypes using terms from the Mammalian Phenotype Ontology and the association of mouse models with human disease. Data release for MGD occur weekly. A summary of the database content for MGD is given in Table 1.

### Enhanced human ortholog detail display

A banner displaying information about the human ortholog of each mouse gene was added to the Gene Detail pages in MGD to improve comparisons of gene-disease associations in mouse and human. The human ortholog detail stripe is positioned above the section of the Gene Detail page that describes alleles and phenotypes for the mouse gene (Figure 1). For each human ortholog, the name and location of the human gene is provided and, if relevant, a list of associated diseases according to the On Line Mendelian Inheritance in Man (OMIM) resource (9) is displayed. The combination of the human ortholog and alleles/phenotypes sections of the Gene Detail page facilitates the ability of the researchers to determine cases where the human gene is associated with a disease and the mouse gene is not (or has yet to be specifically tested as a model) (Figure 1).

By providing information on concordant and discordant instances of mutations in orthologous genes resulting

**Table 1.** Summary of MGD content September 2012

Database content category	September 2012
Number of genes and genome features	37 501
Number of genes with nucleotide sequence data	29 127
Number of genes with protein sequence data	24 738
Number of mouse genes with human orthologs	17 773
Number of mouse genes with rat orthologs	17 253
Number of genes with Gene Ontology (GO) annotations	25 452
Total number of GO annotations	240 562
Number of mutant alleles	712 279
Targeted mutations	47 649
Number of QTL	4715
Number of genotypes with phenotype annotation	44 775
Total number of Mammalian Phenotype (MP) annotations	234 495
Number of mouse models associated to human diseases	3829
Number of references in the MGD bibliography	181 223

in phenotypes that model-specific human diseases MGD can be used to discover potential candidate genes for human diseases that have no gene associations in human; and to discover mutations in mice that should be examined as new models of human disease. For example, the spermatogenesis associated 16 (*Spata16*; MGI: 1918 112) gene is the mouse ortholog for the human SPATA16 (HGNC: 29 935) gene. In humans, mutations in this gene are associated with Apermatogenic Failure 6 (SPGF6) (OMIM 102 530). In mouse, there are currently three alleles for the *Spata16* gene; however, all of these mutants exist only in ES cell lines, thus representing potential sources of mouse models for this disease once the ES cells are made into mice and phenotyped. Conversely, one can observe where a mouse disease model has been associated with a human disease, but there is not yet evidence for the human–disease association to the human ortholog. For example, the mouse cholinergic receptor, muscarinic 3, cardiac (*Chrm3*, MGI: 88398) gene are a model for human Megacystis–Microcolon–Intestinal Hypoperistalsis Syndrome (OMIM 249 210). Thus, study of existing mouse models can facilitate discovery of candidates for disease genes in human.

In some cases, alleles of the mouse gene are associated with human disease phenotypes that differ from associations reported in OMIM. For example, for the mouse caveolin 1 gene (*Cav1*; MGI: 102 709), the human ortholog (CAV1) is associated with congenital lipodystrophy (OMIM: 612 526). However, the genotypes in mouse are associated with human breast cancer (OMIM: 114 480) and Alzheimer's disease (OMIM: 104 300) but not with lipodystrophy. The bicaudal C homolog 1 (*Bicc1*; MGI: 1933 388) gene in mouse is reported as a model for three human diseases in OMIM [Heterotaxy (HTX5), OMIM: 270 100; Polycystic Kidney Disease 1 (PKD1), OMIM: 173 900; and PKD, ARPKD, OMIM: 263 200]. In contrast, the human ortholog, BICC1 (HGNC: 19 351), is not associated with any disease according to OMIM.

### microRNA knockouts

In recent years, the importance of small regulatory RNAs, including microRNAs, in posttranscriptional gene regulation has been recognized. Mice carrying targeted mutations in microRNAs are important resources for characterizing the biological functions of these molecules. Several initiatives have been launched to generate ES cell lines and mice with targeted mutations in microRNAs (10,11). MGD has added these emerging microRNA ‘knockouts’ to the comprehensive catalog of alleles and phenotypes in mouse. Details for microRNA alleles includes the description of the mutation, links to published references, description of observed phenotypes if available and links to the International Mouse Strain Resource (12,13) for information on the availability of strains or cell lines that carry a specific microRNA allele. To date, 434 alleles in 284 microRNAs have been entered into MGD. Although many of these mutant alleles are available as ES cell lines, approximately 170 have been made into live mice. With respect to phenotype

**A**

<b>Human ortholog</b>	<b>SPATA16</b> spermatogenesis associated 16 NCBI Gene ID 83893 Human Synonyms: NYD-SP12, SPGF6 Human Chr3:172607147-172859058 bp, - strand Reference GRCh37.p2 Primary Assembly Human Diseases Associated with Human SPATA16 (1)
<b>Alleles and phenotypes</b>	All alleles(3) : Targeted(2) Gene trapped(1)

**B**

<b>Human ortholog</b>	<b>CAV1</b> caveolin 1, caveolae protein, 22kDa NCBI Gene ID 10530 Human Synonyms: BSCL3, CGL3, MSTP085, VIP21 Human Chr7:116164839-116201239 bp, + strand Reference GRCh37.p2 Primary Assembly Human Diseases Associated with Human CAV1 (1)
<b>Alleles and phenotypes</b>	All alleles(8) : Targeted(8)

**C**

<b>Human ortholog</b>	<b>BICC1</b> bicaudal C homolog 1 (Drosophila) NCBI Gene ID 80114 Human Synonyms: BICC, CYSRD Human Chr10:60272904-60588845 bp, + strand Reference GRCh37.p2 Primary Assembly
<b>Alleles and phenotypes</b>	All alleles(9) : Targeted(2) Spontaneous(1) Chemical(1) Homozygous mutant mice display polycystic kidney disease and die either postnatally or shortly after weaning.

**Human Disease Models Associated with Alleles of Human SPATA16** Close X

**Human Disease**  
[Spermatogenic Failure 6; SPGF6](#)

OMIM ID  
[102530](#)

**Human Disease Models Associated with Alleles of Human CAV1** Close X

**Human Disease**  
[Lipodystrophy, Congenital Generalized, Type 3; CGL3](#)

OMIM ID  
[612526](#)

**Human Disease Models Associated with Alleles of Mouse Cav1** Close X

Diseases listed here are those where a mutant allele of this gene is involved in a mouse genotype used as a model. This does not mean that mutations in this gene contribute to or are causative of the disease.

**Human Disease**  
[Alzheimer Disease; AD](#)  
[Breast Cancer](#)

OMIM ID  
[104300](#)  
[114480](#)

**Human Disease Models Associated with Alleles of Mouse Bicc1** Close X

Diseases listed here are those where a mutant allele of this gene is involved in a mouse genotype used as a model. This does not mean that mutations in this gene contribute to or are causative of the disease.

**Human Disease**  
[Heterotaxy, Visceral, 5, Autosomal; HTX5](#)  
[Polycystic Kidney Disease 1; PKD1](#)  
[Polycystic Kidney Disease, Autosomal Recessive; ARPKD](#)

OMIM ID  
[270100](#)  
[173900](#)  
[263200](#)

**Figure 1.** Screenshots showing the new Human Ortholog and Phenotypic Alleles sections of the MGD Gene Detail page. (A) The SPATA16 gene in humans is associated with a human disease entry in the Online Mendelian Inheritance in Man database, whereas alleles of the orthologous mouse gene have yet to be phenotyped to determine if mutations in the mouse gene result in a good model of the human disease. (B) The CAV1 gene in humans is associated with congenital lipodystrophy, but the available mouse genotypes have not been associated with this disease. Phenotypic alleles of the *Cav1* gene in mouse have been reported for models of human diseases (breast cancer and Alzheimer's disease) that are not associated with the human ortholog. (C) The BICC1 gene in human has not yet been reported as being associated with any human disease, whereas the mouse ortholog has been reported to model several human diseases.

annotations, 67 of the microRNA knockout mice in MGD have phenotype annotations, 5 have no abnormal phenotype and 98 have yet to be phenotyped. As reports appear in the published literature or through large-scale mouse phenotyping projects, the annotations for microRNA knockouts will be updated.

### Phenotype videos

MGD has regularly included still images that illustrate mouse phenotypes associated with alleles and genotypes. Brief video clips of mouse phenotypes have been added recently to provide a new dimension of information on the phenotypic consequences of genomic variants. The over 340 phenotype videos available in MGD are presented as YouTube® clips embedded in the web pages. These videos were generated by the National Heart Lung and Blood Institute's Bench-to-Bassinet program within the Cardiovascular Development Consortium. The imaging modalities represented include Episcope Fluorescence Image Capture (EFIC) image stacks, video microscopy, ultrasound imaging and micro-CT scans. If phenotype

images or videos for alleles of a specific gene are available, a direct link to the images can be found in the Alleles and Phenotypes section of the Gene Detail pages in MGD and on the Phenotype Detail pages for specific mutant alleles. Figure 2 shows a link to the 15 phenotype images associated with alleles of the bicaudal C homolog 1 (*Bicc1*; MGI: 1933388) gene; one of the available images for the *Bicc1*<sup>b2b222C<sup>lo</sup> allele is a 2D serial EFIC image stack of the heart in coronal view. Investigators can submit phenotype videos for either existing or new alleles reported in MGD by following the Submit Data link on the MGI home page and following the instructions for data file submissions.</sup>

### Access to genotype and phenotype data for mouse QTL

MGD staff curate published reports of quantitative trait locus-mapping experiments and, where possible, translates the mapping data into genome coordinates so that regions of the genome associated with mapped phenotypes can be displayed in a genome context. Reciprocal links have been established between quantitative trait loci (QTL) records

<b>Bicc1</b>	
Name	bicaudal C homolog 1 (Drosophila)
ID	MGI:1933888
Synonyms	<i>Bic-C</i> , <i>bgk</i>
Feature Type	protein coding gene
Genetic Map	Chromosome 10 36.75 Mb Consolidated Genetic Map & 1.0b Mapping data[13]
Sequence Map	Chr10:7038550-70622449 bp - strand (From VEGA annotation of NCBI Build 37) <a href="#">VEGA Genome Browser</a>   <a href="#">EntrezGene</a>   <a href="#">Biomart</a>   <a href="#">UCSC Genome Browser</a>   <a href="#">NCBI Map Viewer</a>   <a href="#">NCBI MacView</a>
Mammalian homology	Human: rat: canine: mouse: dog: domestic (Mammalian Orthology) Comparative Map (Mouse/Marmoset/Breca 1 & 2 dm) <a href="#">Gene Tree</a>   <a href="#">Bicc1</a>
Ortholog	<b>Human</b> <a href="#">BICC1</a> bicaudal C homolog 1 (Drosophila) Hecks disease to B0514 Human Chr10:6272904-6298845 bp. + strand Reference: dCh37:5.2 Primary Assembly All alleles[3]   Targeted[2] Spontaneous[1] Chemically induced[5] <b>Alleles</b> : All alleles[3]   Targeted[2] Spontaneous[1] Chemically induced[5] Homozygous mutant mice display polyuria/kidneys with distended abdomens, abnormal bile duct morphology, and die either postnatally or shortly after weaning <b>and phenotypes</b> : Human Diseases Modelled Using Mouse Bicc1 [3]   Alleles Annotated to Human Disease[4] <a href="#">Phenotype Images[15]</a>
<p><b>Caption:</b> 2D serial EFIC image stacks in the co-oral view reveals DORV, common atrium, AVSD and TAPVC  <b>Click thumbnail to play movie.</b> <a href="#">[details]</a></p> <p><b>Represented Alleles:</b>  <code><u>Bicc1</u></code>, <code><u>b2b222Cle</u></code>, bicaudal C homolog 1 (Drosophila); Bench to Bassinet Program (B2B/CvDC), mutation 222 Cecilia Lo</p> <p><b>Copyright:</b> This image is from the Laboratory of Dr. Cecilia Lo, a member of the Cardiovascular Development Consortium (CvDC), Bench to Bassinet (B2B) program of the National Heart Lung and Blood Institute (NHLBI), and is displayed with the permission of the authors. <a href="#">J:175213</a></p>	
<p><b>Caption:</b> Situs inversus totalis with mirror image dextrocardia. <a href="#">[J:1]</a> <a href="#">[details]</a></p> <p><b>Represented Alleles:</b>  <code><u>Bicc1</u></code>, <code><u>b2b222Cle</u></code>, bicaudal C homolog 1 (Drosophila); Bench to Bassinet Program (B2B/CvDC), mutation 222 Cecilia Lo</p> <p><b>Copyright:</b> This image is from the Laboratory of Dr. Cecilia Lo, a member of the Cardiovascular Development Consortium (CvDC), Bench to Bassinet (B2B) program of the National Heart Lung and Blood Institute (NHLBI), and is displayed with the permission of the authors. <a href="#">J:175213</a></p>	
<p><b>Caption:</b> Situs inversus totalis. <a href="#">[details]</a></p> <p><b>Represented Alleles:</b>  <code><u>Bicc1</u></code>, <code><u>b2b222Cle</u></code>, bicaudal C homolog 1 (Drosophila); Bench to Bassinet Program (B2B/CvDC), mutation 222 Cecilia Lo</p> <p><b>Copyright:</b> This image is from the Laboratory of Dr. Cecilia Lo, a member of the Cardiovascular Development Consortium (CvDC), Bench to Bassinet (B2B) program of the National Heart Lung and Blood Institute (NHLBI), and is displayed with the permission of the authors. <a href="#">J:175213</a></p>	

Figure 2. Screenshot showing the link to phenotype images, including a video clip, in the Alleles and Phenotypes section of the Gene Detail page for the *Bicc1* gene.

in MGD records in the QTL Archive (<http://www.qtlarchive.org/>). The QTL Archive extends the utility of mapped phenotypes in MGD by providing researchers with the access to underlying genotype and phenotype data used to map a QTL. Of the 4715 QTL marker records in MGD, over 750 have data available in the QTL Archive.

### Improvements in GO annotation completeness and visualization

MGD is one of the founding members of the Gene Ontology Consortium (GOC) (14,15) and provides major contributions to the development of the GO ontologies and to developing GO community standards for curation of the scientific literature. MGD project curators are responsible for annotating mouse genes and gene products to GO ontology terms.

Improvements in the GO knowledge representation and annotation procedures are incorporated into MGD functional annotation workflows as they are developed (see Gene Ontology Consortium (7). Updated ontologies are loaded into the MGI system and mouse GO annotations are contributed to the GOC on a weekly basis. MGD has the community responsibility of provided non-redundant set of mouse GO annotations to the research community through the MGI database resource and through the

GOC annotation repository and database. The UniProt-Gene Ontology Annotation (GOA) project is the major other provider of mouse GO annotations (16).

New visualization paradigms for displaying GO annotations have been implemented in MGD (Figure 3). The text-based summaries of gene/protein function previously displayed have been replaced by shorter summary statements obtained from NCBI's RefSeq resource (17) for each gene. RefSeq summaries include the source of the summarized information as well as the date the information was last updated. When RefSeq statements are not available for the mouse gene, statements pertaining to the orthologous human gene are included. Orthology assertions between mouse and human genes are taken from NCBI's Homologene resource (18). Previously supported tabular and graphical options for displaying GO annotations are still supported in MGD.

### DATA SUBMISSION

Most of the data in MGD comes from semi-automated curation of the peer-reviewed scientific literature and from collaborative/cooperative arrangements with large, mouse-related data centers and repositories and other informatics resources. MGD also supports electronic data contributions directly from individual researchers.

The screenshot shows the 'Gene Ontology Classifications' page for the gene product of *Kitl*. At the top, there is a table with columns for Symbol, Name, and ID. The symbol is *Kitl*, the name is 'kit ligand', and the ID is 'MGI:96974'. Below this is a section titled 'Go Annotations as Summary Text (Tabular View) (GO Graph)'. It contains a summary from NCBI RefSeq stating: '[Summary is not available for the mouse gene. This summary is for the human ortholog.] This gene encodes the ligand of the tyrosine-kinase receptor encoded by the KIT locus. This ligand is a pleiotropic factor that acts in utero in germ cell and neural cell development, and hematopoiesis, all believed to reflect a role in cell migration. In adults, it functions pleiotropically, while mostly noted for its continued requirement in hematopoiesis. Two transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Jul 2008]'.

Below this is a section titled 'Summary text based on GO annotations supported by experimental evidence in mouse'. It lists several bullet points detailing biological processes, molecular functions, and cellular components supported by experimental evidence. For example, it mentions 'negative regulation of mast cell apoptotic process' and 'positive regulation of cell proliferation'.

At the bottom of the page, there is a note: 'The gene product of *Kitl* has been shown to bind to the gene products of *Kit*. [3]'. There is also a small note about research based on phenotypic analysis of mouse mutants.

Figure 3. Screenshot showing the new Gene Ontology annotation display in MGD.

Any type of data that MGD maintains can be submitted as an electronic contribution. Other common types of submission include mutant and QTL-mapping data. Each electronic submission receives a permanent database accession ID. All data sets are associated with their source, either a publication or an electronic submission reference. MGD reference pages provide links to associated data sets. On-line information about data submission procedures is found at the URL: <http://www.informatics.jax.org/submit.shtml>.

## COMMUNITY OUTREACH AND USER SUPPORT

MGD provides extensive user support through on-line documentation, email and phone access to User Support Staff.

User Support can be accessed by:

- World wide web: <http://www.informatics.jax.org/mgihome/homepages/help.shtml>;
- Email access: [mgi-help@informatics.jax.org](mailto:mgi-help@informatics.jax.org);
- Telephone access: +1 207 288 6445; and
- FAX access: +1 207 288 6132.

Additional outreach and support are provided by a moderated email bulletin board, MGI-LIST (<http://www.informatics.jax.org/mgihome/lists/lists.shtml>). MGI-LIST is managed by the MGI User Support team and has over 2000 subscribers and an average of 75 posts/discussions per month.

## SYSTEM OVERVIEW

The software, database and hardware components comprising MGD are organized into a front end, where the data are made available to the public and a back end, where data are loaded/curated/integrated. Most of the components that were previously supported by a Sybase (<http://www.sybase.com>) relational database management system have been replaced with a combination of PostgreSQL (<http://www.postgresql.org>) and Solr/Lucene indexes (<http://lucene.apache.org/solr>). Solr is an enterprise search server built on the Lucene text searching library. It provides powerful and fast text searching via an applications programming interface (API) over the web (via HTTP). Components maintained outside of the main MGD system include BLAST-able databases and genome assemblies, the databases that support Mouse GBrowse (19) resource and the MGI BioMart (20,21) instance.

There are two primary means by which data are entered into MGD: the editing interface (EI) and automated load programs. The EI is an interactive and graphical application. Curators use the EI to enter new data from the literature, to verify the results of automated loads and to correct errors. The automated load programs integrate larger data sets from many sources into the database. Automated loads involve quality control checks and processing algorithms that integrate the bulk of the data automatically and identify issues to be resolved by curators or the data provider. Through these two vehicles,

the EI and automated loads, MGD is able to scale and adapt as new data sources for the mouse are made available.

Access to information in MGD is provided in several ways to support our diverse community of users including the web interface, Batch Query tool, FTP and a web services API.

## Web interface

Interactive web-based interfaces are the primary means of access to MGD. The keyword based 'Quick Search' option on the MGI home page is the most commonly used search tool for single concept searches. The *Batch Query* tool (19) is a component of the MGD web interface that enables searches by lists of genes. It can be used as an accession ID translator (e.g. to convert a list of MGI IDs to the corresponding list of EntrezGene IDs) or as a way to retrieve a set of information for a collection of genes/features (e.g. to obtain all the GO annotations for a list of genes). In either case, the input is a user-specified collection of IDs in a text field or file upload. Gene symbols and a wide variety of ID types are accepted, including IDs from MGI, EntrezGene, Ensembl, Havana/Vega, GenBank, RefSeq, UniProt, RefSNP, Affymetrix, GO, etc. Users also select their desired output, including annotations from GO, MP, OMIM; phenotypic alleles; gene expression results (from GXD); or any of the above ID types. The Batch Query maps each input ID to any corresponding genes/features in MGD (may be more than one) and returns them with the requested data. The Batch Query is fully integrated with the MGD web interface and is called from various pages to generate a user customizable gene/feature summary. Results are available as HTML, tab delimited or Excel format.

Other web interfaces to MGD include MouseBLAST for sequence similarity searches against a variety of rodent-relevant sequence databases, Mouse GBrowse for genome centric browsing and MGI's BioMart for searches that combine results from MGI and Ensembl.

*MGD's public FTP reports* include over 50 flat file reports that are generated weekly. Most external informatics resources that incorporate data from MGD obtain their data from these reports. Custom reports are created upon request.

The MGI web services API is a Simple Object Access Protocol-based interface to the database providing programmatic access with identical functionality as the Batch Query tool described above.

## CITING MGD

For a general citation of the MGD resource, researchers should cite this article. In addition, the following citation format is suggested when referring to data sets specific to the MGD component of MGI: MGD, MGI, The Jackson Laboratory, Bar Harbor, Maine (URL: <http://www.informatics.jax.org>). [Type in date (month, year) when you retrieved the data cited.]

## ACKNOWLEDGEMENTS

The Mouse Genome Database Group: M.T. Airey, A. Anagnostopoulos, R. Babik, R.M. Baldarelli, J.S. Beal, S.M. Bello, N.E. Butler, J. Campbell, L.E. Corbani, H. Dene, H.R. Drabkin, K.L. Forthofer, S.L. Giannatto, M. Knowlton, J.R. Lewis, M. McAndrews, S. McClatchy, D.S. Miers, L. Ni, H. Onda, J.E. Ormsby, J.M. Recla, D.J. Reed, B. Richards-Smith, D.R. Shaw, D. Sitnikov, C.L. Smith, M. Tomczuk, L.L. Washburn, Y. Zhu.

## FUNDING

National Institutes of Health; National Human Genome Research Institute [HG000330]. Funding for open access charge: Grant funds.

*Conflict of interest statement.* None declared.

## REFERENCES

- Bradley,A., Anastassiadis,K., Ayadi,A., Battey,J.F., Bell,C., Birling,M.C., Bottomley,J., Brown,S.D., Burger,A., Bult,C.J. *et al.* (2012) The mammalian gene function resource: the international knockout mouse consortium. *Mamm. Genome*, **23**, 580–6.
- Brown,S.D. and Moore,M.W. (2012) Towards an encyclopaedia of mammalian gene function: the International Mouse Phenotyping Consortium. *Dis. Model Mech.*, **5**, 289–292.
- Threadgill,D.W. and Churchill,G.A. (2012) Ten years of the collaborative cross. *G3*, **2**, 153–156.
- Churchill,G.A., Gatti,D.M., Munger,S.C. and Svenson,K.L. (2012) The diversity outbred mouse population. *Mamm. Genome*, **23**, 713–8.
- Finger,J.H., Smith,C.M., Hayamizu,T.F., McCright,I.J., Eppig,J.T., Kadin,J.A., Richardson,J.E. and Ringwald,M. (2011) The mouse Gene Expression Database (GXD): 2011 update. *Nucleic Acids Res.*, **39**, D835–D841.
- Begley,D.A., Krupke,D.M., Neuhauser,S.B., Richardson,J.E., Bult,C.J., Eppig,J.T. and Sundberg,J.P. (2012) The Mouse Tumor Biology Database (MTB): a central electronic resource for locating and integrating mouse tumor pathology data. *Vet. Pathol.*, **49**, 218–223.
- Gene Ontology Consortium. (2012) The Gene Ontology: enhancements for 2011. *Nucleic Acids Res.*, **40**, D559–D564.
- Evsikov,A.V., Dolan,M.E., Genrich,M.P., Patek,E. and Bult,C.J. (2009) MouseCyc: a curated biochemical pathways database for the laboratory mouse. *Genome Biol.*, **10**, R84.
- Amberger,J., Bocchini,C. and Hamosh,A. (2011) A new face and new challenges for Online Mendelian Inheritance in Man (OMIM(R)). *Hum. Mutat.*, **32**, 564–567.
- Prosser,H.M., Koike-Yusa,H., Cooper,J.D., Law,F.C. and Bradley,A. (2011) A resource of vectors and ES cells for targeted deletion of microRNAs in mice. *Nat. Biotechnol.*, **29**, 840–845.
- Park,C.Y., Jeker,L.T., Carver-Moore,K., Oh,A., Liu,H.J., Cameron,R., Richards,H., Li,Z., Adler,D., Yoshinaga,Y. *et al.* (2012) A resource for the conditional ablation of microRNAs in the mouse. *Cell Rep.*, **1**, 385–391.
- Eppig,J.T. and Strivens,M. (1999) Finding a mouse: the International Mouse Strain Resource (IMSR). *Trends Genet.*, **15**, 81–82.
- Eppig,J.T., Bult,C.J., Kadin,J.A., Richardson,J.E., Blake,J.A., Anagnostopoulos,A., Baldarelli,R.M., Baya,M., Beal,J.S., Bello,S.M. *et al.* (2005) The Mouse Genome Database (MGD): from genes to mice—a community resource for mouse biology. *Nucleic Acids Res.*, **33**, D471–D475.
- Blake,J.A. and Harris,M.A. (2008) The Gene Ontology (GO) project: structured vocabularies for molecular biology and their application to genome and expression analysis. *Curr. Protoc. Bioinformatics*, Chapter 7, Unit 7.2.
- Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
- Dimmer,E.C., Huntley,R.P., Alam-Faruque,Y., Sawford,T., O'Donovan,C., Martin,M.J., Bely,B., Browne,P., Mun Chan,W., Eberhardt,R. *et al.* (2012) The UniProt-GO Annotation database in 2011. *Nucleic Acids Res.*, **40**, D565–D570.
- Pruitt,K.D., Tatusova,T., Brown,G.R. and Maglott,D.R. (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.*, **40**, D130–D135.
- Jenuth,J.P. (2000) The NCBI: Publicly available tools and resources on the Web. *Methods Mol. Biol.*, **132**, 301–312.
- Bult,C.J., Eppig,J.T., Kadin,J.A., Richardson,J.E. and Blake,J.A. (2008) The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res.*, **36**, D724–D728.
- Smedley,D., Haider,S., Ballester,B., Holland,R., London,D., Thorisson,G. and Kasprzyk,A. (2009) BioMart—biological queries made easy. *BMC Genomics*, **10**, 22.
- Bult,C.J., Kadin,J.A., Richardson,J.E., Blake,J.A. and Eppig,J.T. (2010) The Mouse Genome Database: enhancements and updates. *Nucleic Acids Res.*, **38**, D586–D592.